

General Disclaimer

One or more of the Following Statements may affect this Document

- This document has been reproduced from the best copy furnished by the organizational source. It is being released in the interest of making available as much information as possible.
- This document may contain data, which exceeds the sheet parameters. It was furnished in this condition by the organizational source and is the best copy available.
- This document may contain tone-on-tone or color graphs, charts and/or pictures, which have been reproduced in black and white.
- This document is paginated as submitted by the original source.
- Portions of this document are not fully legible due to the historical nature of some of the material. However, it is the best reproduction available from the original submission.

(NASA-CR-138609) ON THE ERROR
PROBABILITY OF GENERAL TREE AND TRELLIS
CODES WITH APPLICATIONS TO SEQUENTIAL
DECODING (Notre Dame Univ.) 34 p HC
\$4.75

N74-27069

Unclas

CSCL 12A G3/19 41311

Department of

ELECTRICAL ENGINEERING



UNIVERSITY OF NOTRE DAME, NOTRE DAME, INDIANA

ON THE ERROR PROBABILITY OF
GENERAL TREE AND TRELLIS CODES WITH
APPLICATIONS TO SEQUENTIAL DECODING*

Rolf Johannesson
Dept. of Electrical Engineering
University of Notre Dame
Notre Dame, Indiana 46556

Technical Report No. EE-7316
December 1973

* This work was supported in part by the National Aeronautics and Space Administration under NASA Grant NGL 15-004-026 at the University of Notre Dame in liaison with the Communications and Navigation Division of the Goddard Space Flight Center, and in part by the Laboratory for Telecommunication Theory, The Lund Institute of Technology, Lund, Sweden.

On the Error Probability of General Tree and Trellis Codes with
Applications to Sequential Decoding.¹⁾

by

Rolf Johannesson

(Submitted to IEEE Transactions on Communications)

ABSTRACT - An upper bound on the average error probability for maximum-likelihood decoding of the ensemble of random binary tree codes is derived and shown to be independent of the length of the tree.

An upper bound on the average error probability for maximum-likelihood decoding of the ensemble of random L-branch binary trellis codes of rate $R = 1/n$ is derived which separates the effects of the tail length T and the memory length M of the code. It is shown that the bound is independent of the length L of the information sequence when $M \geq T + [nE_{VU}(R)]^{-1} \log_2 L$. This implication is investigated by computer simulations of sequential decoding utilizing the stack algorithm. These simulations confirm the implication and further suggest the following empirical formula for the true undetected decoding error probability with sequential decoding:

$$P[\epsilon] \approx c 2^{-nTR_T}, \text{ provided } M \geq T + [nE_{VU}(R)]^{-1} \log_2 L$$

where c is a constant independent of L , T and M . The exponent R_T is related to Viterbi's upper and lower exponents for the ensemble of time-varying convolutional codes by the inequality:

$$E_{VU}(R) \leq R_T \leq E_{VL}(R).$$

1) The first draft of this paper was presented orally at the IEEE International Symposium on Information Theory, Ashkelon, Israel, June 25-29, 1973.

On the Error Probability of General Tree and Trellis Codes with
Applications to Sequential Decoding.¹⁾

by

Rolf Johannesson

(Submitted to IEEE Transactions on Communications)

ABSTRACT - An upper bound on the average error probability for maximum-likelihood decoding of the ensemble of random binary tree codes is derived and shown to be independent of the length of the tree.

An upper bound on the average error probability for maximum-likelihood decoding of the ensemble of random L-branch binary trellis codes of rate $R = 1/n$ is derived which separates the effects of the tail length T and the memory length M of the code. It is shown that the bound is independent of the length L of the information sequence when $M \geq T + [nE_{VU}(R)]^{-1} \log_2 L$. This implication is investigated by computer simulations of sequential decoding utilizing the stack algorithm. These simulations confirm the implication and further suggest the following empirical formula for the true undetected decoding error probability with sequential decoding:

$$P[\epsilon] \approx c 2^{-nR_T T}, \text{ provided } M \geq T + [nE_{VU}(R)]^{-1} \log_2 L$$

where c is a constant independent of L , T and M . The exponent R_T is related to Viterbi's upper and lower exponents for the ensemble of time-varying convolutional codes by the inequality:

$$E_{VU}(R) \leq R_T \leq E_{VL}(R).$$

1) The first draft of this paper was presented orally at the IEEE International Symposium on Information Theory, Ashkelon, Israel, June 25-29, 1973.

I. INTRODUCTION

Massey [1] has recently defined the class of random tree codes as a generalization of the type of convolutional code used in sequential decoding and has defined the class of random trellis codes as a generalization of the type of convolutional code used in Viterbi decoding. He has also proved random upper bounds on the average probability of error for maximum-likelihood decoding of these codes for codes rates less than R_0 , where R_0 is less than capacity C of the channel.

In Section III of this paper, we extend Massey's bound for tree codes to all rates less than capacity. In Section IV we do the same for his bound on trellis codes but we do this in the context of a more general class of trellis codes for which a distinction can be made between its "memory length" and its "tail length". The bounds obtained suggest that it is advantageous to use a memory length which is a specified amount greater than the tail length, this amount depending on the length of the trellis. In Section V, we report sequential decoding simulations which confirm this suggestion and which should be useful guides in the design of future sequential decoding systems. These simulations suggest an empirical formula for the true decoding error probability with sequential decoding which we give in Section VI.

To relate the more general codes used in this paper to the special cases used in practice, we note that convolutional codes constitute the class of linear tree codes. After L information bits have been encoded in the convolutional code, the encoded sequence is terminated after a "tail" of T information zeroes has been encoded. A convolutional code is further characterized by its encoding memory length M which is the number of unit delays in the encoder. The usual practice has been to take $M=T$ but the merit of removing this restriction will become obvious in the sequel.

II. PRELIMINARIES

In our discussion of tree and trellis codes, we shall restrict our attention to codes of rate $R = 1/n$ both for simplicity of description and because these rates are those of the most practical interest.

A binary tree code of rate $R = 1/n$, tree length L , and tail length T is formed by assigning n channel input symbols to each branch of a rooted tree such that 2 branches stem from the root node, 2 branches stem from each successive node at depth i from the root for $i < L$, and a single branch stems from each node at depth i from the root for $L \leq i < L+T$. We show such a tree code in Figure 1 for $T=3$, $n=2$, and a binary input channel. To encode a binary information sequence of length L with such a code, one begins at the root node and moves through the tree taking the upper branch or the lower branch according as each successive information digit is 0 or 1. Since $n(L+T)$ encoded digits result from this process, the true rate (in bits per channel input symbol) of the code is $L/[n(L+T)]$ and hence approximately equal to its supremum $R = 1/n$ only when $L \gg T$. For this reason, one ordinarily desires to choose $L \gg T$. We shall call those nodes where the tree divides "information nodes" to correspond with their function in encoding the binary information sequence of length L .

In this paper we shall derive an upper bound on the average decoding error probability, assuming maximum-likelihood decoding, for the ensemble of R, L, T tree codes in which each channel input symbol in the tree is chosen independently according to a specified probability distribution.

III. UPPER BOUND FOR RANDOM TREE CODES.

For the binary tree codes described in II, we shall now derive an upper bound on the average probability of decoding error, assuming that a maximum-likelihood decoder is used.

Suppose that L information bits have been encoded and that the resultant $n(L+T)$ encoded symbols have been transmitted through a discrete memoryless channel. Let E_i ($i = 1, 2, \dots, L$) be the event that the probability of the received sequence given some incorrect path stemming from the i -th last information node along the correct path is equal to or greater than the probability of the received sequence given the correct path. Then, letting ϵ be the event that a decoding error is made by the maximum-likelihood decoder, we have

$$\epsilon \subset E_1 \cup E_2 \cup \dots \cup E_L \quad (1)$$

(where we are prevented from writing " $=$ " rather than " \subset " only by the fact that the maximum-likelihood decoder may correctly decode in case of ties for the best path). The average probability of error over the ensemble of R, L, T tree codes is then bounded by the union bound

$$\overline{P[\epsilon]} \leq \overline{P[E_1]} + \overline{P[E_2]} + \dots + \overline{P[E_L]} \quad (2)$$

We now note that a total of 2^{i-1} incorrect paths, each of length $T+i$ branches, stem from the i -th last information node on the correct path. Since over the ensemble of codes all symbols on each path are mutually independent, the symbols on the transmitted path stemming from the i -th last node are independent of the symbols on any incorrect path stemming from the same node. Thus, we may use Gallager's random coding upper bound on block codes of length $n(T+i)$ with $(2^{i-1}+1)$ codewords [2] to bound the i -th term on the righthand side of (2) as

$$\overline{P[E_i]} \leq 2^{(i-1)\rho} 2^{-n(T+i)E_0(\rho)}, \quad 0 \leq \rho \leq 1 \quad (3)$$

where ρ is a parameter which can be chosen later. The function $E_0(\rho)$ is defined as

$$E_0(\rho) = \max_{Q(x)} \left\{ -\log_2 \sum_Y \left[\sum_X Q(x) p(y|x)^{1/(1+\rho)} \right]^{1+\rho} \right\} \quad (4)$$

where X and Y are the channel input and output spaces respectively, $p(y|x)$ is the channel transition probability distribution, and $Q(x)$ is a probability distribution on the input space. The maximizing $Q(x)$ in (4) is then also taken as the specific probability distribution defining the ensemble of R, L, T tree codes.

Inserting (3) in (2) yields

$$\begin{aligned} \overline{P[\epsilon]} &\leq \sum_{i=1}^L 2^{(i-1)\rho} 2^{-n(T+i)E_0(\rho)} \\ &< 2^{-nTE_0(\rho)} 2^{-\rho} \sum_{i=1}^{\infty} 2^{-i[nE_0(\rho) - \rho]} \\ &= \frac{2^{-nE_0(\rho)}}{1 - 2^{-n\delta}} \cdot 2^{-nTE_0(\rho)}, \quad 0 \leq \rho \leq 1 \end{aligned} \quad (5)$$

where

$$\delta = E_0(\rho) - \rho R > 0 \quad (6)$$

Since $E_0(\rho)$ is a monotonically increasing function of ρ , we have

Theorem 1:

The average probability of error for maximum-likelihood decoding of the ensemble of binary $R = 1/n, L, T$ tree codes satisfies

$$\overline{P[\epsilon]} < c 2^{-nTE_{VU}(R)} \quad (7)$$

where $E_{VU}(R)$ is Viterbi's upper bound exponent [3], namely

$$E_{VU}(R) = \begin{cases} R_0 & 0 \leq R < R_0 \\ \sup_{\rho} E_0(\rho) & R_0 \leq R < C \end{cases} \quad (8)$$

where the supremum is taken over ρ such that $0 \leq \rho \leq 1$ and $E_0(\rho) > \rho R$ and where

$$c = \frac{2^{-nE_{VU}(R)}}{1 - 2^{-n\delta}} \quad (9)$$

The "constant" c depends on R but is independent of L and T . The exponent R_0 is defined to be $E_0(1)$. This exponent R_0 is numerically equal to the "computational cutoff rate R_{comp} [4] encountered in sequential decoding. The rate C is the channel capacity.

The remarkable feature of the bound (7) is its independence of the length L of the tree! The bound (7) implies that only the tail length T is important in determining the error probability for tree codes. In Section V, we report simulations which verify this conclusion.

Viterbi [5] has given the same upper bound on the first-event error probability for time-varying convolutional codes.

We also wish to remark that for rates $R < R_0$, Massey [1] derived an upper bound equal to (7). His argument used the two codeword exponent R_0 and this work stimulated the investigation reported in this paper to extend the bound from R_0 to C and to conduct simulations to verify its implications.

Massey [6] also recently presented the straightforward generalization of his formula to rates $R = k/n < R_0$, where k and n are integers and suggested the use in principle of convolutional codes with memory length greater than tail length to remove the dependence of $P[\epsilon]$ on tree length L as has been carried out in the simulations reported in

Section V. Massey's generalized argument can be extended to show that (7) holds for all rates $R = k/n$ provided that the constant "c" in (7) is given by

$$c = (2^k - 1)^p \cdot \frac{2^{-nE_{VJ}(R)}}{1 - 2^{-n\delta}} \quad (10)$$

rather than by (9). We have omitted this refinement because the generalization while straightforward is somewhat awkward and because $R = 1/n$ is the case of greatest practical interest.

IV. UPPER BOUNDS FOR TRELLIS CODES.

In Section II, we described an $R = 1/n$, L , T tree code as the assignment of $n = 1/R$ channel input symbols to each branch of a particular rooted tree. Recall that a sequence of L information bits specifies precisely one path through this tree. We now define an R , L , T , M trellis code to be a tree code with the property that

- (i) if the preceeding M information digits on the paths leading into two nodes at the same depth in the tree coincide then the same further encoded sequence results whenever the same further information sequence is applied starting from either node and
- (ii) all digits are the same on the last $M-T$ branches.

In other words, the "memory" or dependence on the past information bits is limited to the M previous information bits but the useful "tail" of the tree is only T rather than M branches in length. From an encoding viewpoint then, nodes with same preceeding M information bits can be "merged" in the tree so that the possible encoding paths may be shown as forming a "trellis-like structure". In Figure 2, we show an $R = 1/2$ binary trellis code with $T=1$ and $M=2$ for a binary input channel. Forney [4] was the first to use the term "trellis" in connection with a special class of such codes (viz. convolutional codes with $M=T$) while Massey [1] generalized the definition to that given here except again for the restriction that $M=T$. By allowing $T < M$ we are able, as shown in the sequel, to demarcate rather precisely the different effects of the "tail length" T and the "memory length" M on decoding error probability.

Our artifice of requiring all of the digits on the last $M-T$ branches of each path in the trellis to coincide renders these digits "useless" and hence unnecessary to transmit over the discrete memoryless channel being considered and hence to have a true "tail" of length only T branches, but this artifice also allows us to use with only slight change the bounding techniques normally used for the "usual" trellis codes with $M=T$.

Suppose now that L information bits have been encoded in an R, L, T, M trellis code and that first $n(L+T)$ encoded digits on the corresponding path (i.e. all except the "useless" last $n(M-T)$ digits which are the same on all paths in the trellis) and $n(L+T)$ corresponding digits have been received over the discrete memoryless channel.

Consider next any subpath of the correct path in the trellis. We define an "adversary" for this subpath to be any path which has the same first node and "remerge" with this subpath at its last node, i.e. it has this same last node but no previous node in common with this subpath (except of course the first node). By our definition of a trellis code, an adversary must have length at least $M+1$ branches since after diverging with the correct path at some node there must be some "consecutive information bits that agree with those on the correct path for remergence to take place.

A maximum-likelihood decoder for the trellis code will decode correctly unless there is some subpath of the correct path such that the probability of the corresponding portion of the received sequence given some adversary of this subpath is as great or greater than its probability given the subpath. In case of ties for the best subpath the maximum-likelihood decoder may decode correctly. Hence we begin our bounding of the decoding error probability by defining F_j ($M < j \leq L+M$) as the event that for some subpath of the correct path ending at the j -th node from the root along the correct path the corresponding portion of the received sequence is as probable or more probable given some adversary of the subpath than given the subpath. Letting ϵ be the event that the decoding is not correct, we then have

$$\epsilon \subset F_{M+1} \cup F_{M+2} \cup \dots \cup F_{M+L} \quad (11)$$

Using the union bound, we can then overbound the average error probability for the ensemble of trellis codes in which each digit in the trellis is chosen independently according to some probability distribution $Q(\cdot)$ over the channel input space as

$$P[\epsilon] \leq P[F_{M+1}] + P[F_{M+2}] + \dots + P[F_{M+L}] \quad (12)$$

and we can overbound $\overline{P[F_j]}$ using the random coding bound for certain ensembles of block codes as we now consider in detail. For the node at depth j from the root along the correct path, there is only one adversary of a subpath of length $M+1$ branches which remerges at this node. For $\ell \geq 2$, there are at most $2^{\ell-2}$ adversaries of a subpath of length $M+\ell$ which remerges at this node as can be seen from the fact that the first information bit for the adversary must disagree with that of the subpath while the last M information bits must agree and the information bit just previous to these must disagree with those of the subpath or remergence would have occurred sooner. Hence $M+2$ of the $M+\ell$ information bits of an adversary of length $M+\ell$ are uniquely specified when $\ell \geq 2$. Thus $2^{\max(0, \ell-2)}$ is a general upper bound on the number of adversaries for a subpath of length $M+\ell$ branches, $\ell \geq 1$.

In considering nodes at depth j from the root, for $M < j \leq L+T$ all digits on the adversaries remerging at this node are statistically independent of those on the corresponding correct subpath so that using Gallager's upper bound on random block codes [2] we have

$$\begin{aligned}
 \overline{P[F_j]} &< \sum_{\ell=1}^{j-M} (2^{\max(0, \ell-2)})^{\rho} 2^{-n(M+\ell)E_0(\rho)} \\
 &< (2^{-nE_0(\rho)} + 2^{-2\rho} \sum_{\ell=2}^{\infty} 2^{-n\delta\ell}) 2^{-nME_0(\rho)} \\
 &< \frac{2^{-nE_0(\rho)}}{1-2^{-n\delta}} \cdot 2^{-nME_0(\rho)} \\
 &= c 2^{-nME_0(\rho)}, \quad 0 \leq \rho \leq 1, \quad M < j \leq L+T
 \end{aligned} \tag{13}$$

But for $L+T < j \leq L+M$, the digits on the last $(j-M+T)$ branches of each adversary agree with those on the correct path because of our artifice of using the same channel input letter for all digits on the last $M-T$ branches of every path in the trellis. Thus the block coding bound must be revised to account for the reduced useful codeword length and we obtain

$$\begin{aligned}
\overline{P[F_j]} &< \sum_{\ell=1}^{j-M} (2^{\max(0, \ell-2)}) \rho \cdot 2^{-n(L+M-j+T+\ell)} E_0(\rho) \\
&< 2^{-n(L+M-j)} E_0(\rho) (2^{-nE_0(\rho)} \\
&\quad + 2^{-2\rho} \sum_{\ell=2}^{\infty} 2^{-n\delta\ell}) 2^{-nTE_0(\rho)} \\
&< 2^{-n(L+M-j)} E_0(\rho) c 2^{-nTE_0(\rho)}, \\
0 \leq \rho \leq 1, \quad L+T < j \leq L+M
\end{aligned} \tag{14}$$

where in both (13) and (14)

$$\delta = E_0(\rho) - \rho R > 0 \tag{15}$$

and where c is as given in (9), i.e. c is the same constant as in (7).

Finally, substituting (13) and (14) in (12) we have

$$\begin{aligned}
\overline{P[\epsilon]} &< \sum_{j=M+1}^{L+T} c 2^{-nME_0(\rho)} \\
&\quad + \sum_{j=L+T+1}^{L+M} 2^{-n(L+M-j)} E_0(\rho) c 2^{-nTE_0(\rho)} \\
&= (L+T-M) c 2^{-nME_0(\rho)} + \frac{1-2^{-n(M-T)} E_0(\rho)}{1-2^{-nE_0(\rho)}} c 2^{-nTE_0(\rho)}, \\
0 \leq \rho \leq 1, \quad T \leq M
\end{aligned} \tag{16}$$

Since $E_0(\rho)$ is a monotonically increasing function of ρ , we have

Theorem 2:

The average probability of error for maximum-likelihood decoding of the ensemble of binary $R = 1/n$, L , T , M trellis codes satisfies

$$\overline{P[\epsilon]} < c 2^{-nTE_{VU}(R)} \left[\frac{1-2^{-n(M-T)E_{VU}(R)}}{1-2^{-nE_{VU}(R)}} + (L+T-M)2^{-n(M-T)E_{VU}(R)} \right] \quad (17)$$

where $E_{VU}(R)$ is given in (8) and $T \leq M$.

Upon observing that, since $T \leq M$,

$$\frac{1-2^{-n(M-T)E_{VU}(R)}}{1-2^{-nE_{VU}(R)}} \leq M-T \quad (18)$$

we can state

Corollary 1:

The average probability of error for maximum-likelihood decoding of the ensemble of binary $R = 1/n$, L, T, M codes satisfies

$$\overline{P[\epsilon]} < Lc 2^{-nTE_{VU}(R)} \quad (19)$$

where $E_{VU}(R)$ is given in (8). In the special case when $T=M$, the bound of Corollary 1 is identical to Viterbi's well-known upper bound for the ensemble of time-varying convolutional codes.

Next, we notice that the first term within the brackets in (17) is independent of L whereas the second term can be made arbitrarily small for a given L by increasing M . Thus, by choosing that value of M which, for a given L , makes these two terms equal we have

Corollary 2:

The average probability of error for maximum-likelihood decoding of the ensemble of binary $R = 1/n$, L, T, M codes satisfies

$$\overline{P[\epsilon]} < c' 2^{-nTE_{VU}(R)} \quad (20)$$

provided

$$M \geq T + [nE_{VU}(R)]^{-1} \log_2 L \quad (21)$$

where

$$c' = \frac{2c}{1 - 2^{-nE_{VU}(R)}} \quad (22)$$

and $E_{VU}(R)$ is given in (8).

The bound (20) of Corollary 2 is independent of the length L of the trellis and is very similar to the bound (7) of Theorem 1 for the ensemble of tree codes.

We remark that Theorem 2 and its corollaries can be proved to hold for the ensemble of time-varying convolutional codes, but not presently for the ensemble of constant convolutional codes which lack the independence needed in the proof. However, we conjecture that Theorem 2 and its corollaries hold also for the ensemble of constant convolutional codes which in fact are the type of convolutional code that has always been used in practice. Since there must always be at least one code whose $P[\epsilon]$ is no more than average, we can state an even weaker

Conjecture:

The probability of error for maximum-likelihood decoding of a "good" binary $R = 1/n$, L , T , M constant convolutional code satisfies

$$P[\epsilon] \leq c 2^{-nTE_{VU}(R)} \quad (23)$$

provided

$$M \geq T + [nE_{VU}(R)]^{-1} \log_2 L \quad (24)$$

where c is a constant independent of L , T and M and $E_{VU}(R)$ is given in (8).

The conjecture is given strong support by the simulations discussed in Section V which is not too surprising since all "Viterbi type" error bounds for convolutional codes can presently be proved only for random or time-varying codes but all simulations to date have used "good" constant codes and the bounds have always been found to be valid, i.e. the actual $P[e]$ for the constant code considered was smaller than the upper bound on $P[e]$ for the ensemble of time-varying codes of that length.

Finally, we note that by taking $M = L+T$, the ensemble of R, L, T, M trellis codes becomes exactly the ensemble of R, L, T tree codes. We have already noted that for $M=T$, the ensemble of R, L, T, M trellis codes becomes the ensemble of trellis codes defined by Massey [1]. Hence our Theorem 2 is a generalization from which upper bounds on $P[e]$ for both these ensembles follow as special cases.

V. RESULTS OF SIMULATIONS.

In order to test the implications of the bounds for trellis codes derived in the previous section and in particular to test our conjecture that these bounds apply to "good" constant convolutional codes, decoding simulations for the binary symmetric channel (BSC) were conducted.

Although the theory was developed for true maximum-likelihood decoding, it is well-known [7] that the exponent of error probability for sequential decoding is the same as that for true maximum-likelihood or "Viterbi" decoding. Since the latter is too time-consuming for practical simulations except when M is very small, it was decided to perform the simulations using sequential decoding. The particular sequential decoding algorithm employed was the quantized or "stack bucket" algorithm proposed by Jelinek [8] which is the practical modification of the "stack algorithm" conceived independently by Zigangirov [9] and Jelinek. The simulations were all performed for the code rate $R = 1/2$. The "good" convolutional codes chosen were the "complementary codes" found by Bahl and Jelinek [10]. Three different BSC's were simulated, namely those with "crossover probability" p of 0.033, 0.045 and 0.057 which correspond to $R = 0.9 R_0$, $R = R_0$ and $R = 1.1 R_0$ respectively when $R = 1/2$. For each code used on each of these channels, a very large number (up to 60,000) of received "frames", i.e. complete received sequences of length $n(L+T)$, were decoded so that the decoding error probability could be accurately inferred. The "metrics" used for the sequential decoding on each BSC are tabulated in Table I.

In Figures 3, 4 and 5, we give the simulation results for the sequential decoding undetected error probability $P[\epsilon]$ as a function of the tail length T of the convolutional code. (Because of the extreme variability of the computation in sequential decoding when M is large, there were occasions where the decoding had to be stopped because the computation exceeded the allotted maximum. The observed probability of this "overflow" is tabulated in Table II and had negligible effect on the curves of Figures 3, 4 and 5.) These curves show that the actual $P[\epsilon]$ decreases exponential with T with an exponent very close to that of the bound (20) for the range $T < M - [nE_{VU}(R)]^{-1} \log_2 L + 2$

while further increases in T beyond this point have virtually no effect on $P[\epsilon]$. This is in surprising agreement with the effect of M and T on $P[\epsilon]$ in the bound of (17). It is rather remarkable that the range of T for which the bound becomes independent of L , viz. $T \leq M - [nE_{VU}(R)]^{-1} \log_2 L$ is so close to the range where the true $P[\epsilon]$ becomes independent of L . Hence the relation (21) can be taken as a slightly conservative design rule for choosing M so that $P[\epsilon]$ is reduced to as little as possible for the tail length T that can be allocated to an encoded frame.

VI. AN EMPIRICAL FORMULA FOR $P[\epsilon]$ WITH SEQUENTIAL DECODING.

The curves of Figures 3, 4 and 5 for $P[\epsilon]$ versus T obtained from our sequential decoding simulations are well approximated by two straight lines giving the exponential decrease of $P[\epsilon]$ with T up to the point where $P[\epsilon]$ becomes independent of T . The following empirical formula then provides a close match to the undetected error probability for sequential decoding of the "good" convolutional codes used in these simulations:

$$P[\epsilon] \approx c 2^{-n T R_T}, \text{ provided } M \geq T + [n E_{VU}(R)]^{-1} \log_2 L \quad (25)$$

where the observed values of c and R_T for each of the decoding simulations performed are given in Tables III, IV and V. The near constancy of these parameters for wide variations in M and L when $M \geq 6$ suggest that these parameters can be well estimated in advance and used for design of sequential decoding systems. The case $M=4$ is a case where the memory length is so small that the exponential approximation is not very well fulfilled. In fact, the apparent slight variation of c and R_T for large values of L is probably related more to the inaccuracies of the statistical values because of the small but increasingly non-negligible overflow probability P_0 (as given in Table II) rather than to an actual variation of c and R_T .

The average values of R_T evaluated over $M \geq 8$ and over all L (four values exceeding 0.96 are omitted) are given together with Viterbi's upper and lower exponents $E_{VU}(R)$ and $E_{VL}(R)$ for $R = 1/2$ in Table VI. The exponents $E_{VU}(R)$ and $E_{VL}(R)$ are shown in Figure 6 where straight-line approximations are used when $R_0 < R < C$.

From Table VI we conclude that

$$E_{VU}(R) \leq R_T \leq E_{VL}(R), \quad R < R_0 \quad (26)$$

$$R_T = E_{VU}(R) = E_{VL}(R), \quad R_0 \leq R < C \quad (27)$$

Thus, R_T is in agreement with both the exponent $E_{VU}(R)$ of the upper bound (20) and the exponent $E_{VL}(R)$ of Viterbi's lower bound on the error probability in decoding a time-varying convolutional code [3].

VII. REMARK.

Finally, we should remark that, if we wanted solely to minimize the undetected error probability with sequential decoding for a given memory length and was not concerned with holding the tail size to a minimum to maximize the true rate of the trellis code then the optimal value of the tail length is, of course, the memory length, i.e. $T=M$. Probably this fact has caused investigators to ignore the distinction between the "tail" and the "memory" so that the memory length came to be honoured for work actually done by the tail.

VIII. ACKNOWLEDGEMENT.

The author is greatly indebted to professor James L. Massey, for his help in presenting these results in an easily understandable manner. Professor Göran Einarsson is acknowledged for his comments upon the first draft of this paper.

REFERENCES

- [1] J.L. Massey
"Coded Digital Communication".
Notes used in course 4007 - fall 1971 at the
Royal Technical University of Denmark, Lyngby, Denmark
- [2] R.G. Gallager
"Information Theory and Reliable Communication".
New York: Wiley, 1968, p 138
- [3] A.J. Viterbi
"Error bounds for convolutional codes and an asymptotically optimum
decoding algorithm".
IEEE Trans. Inform. Theory, IT-13, April 1967, 260-269
- [4] G.D. Forney, Jr.
"Final report on a coding system design for advanced solar missions".
Codex Corp., Watertown, Mass., December 1967, appendix A.
- [5] A.J. Viterbi
"Convolutional codes and their performance in communication systems".
IEEE Trans. Commun. Technology, vol COM-19, October 1971, 751-772
- [6] J.L. Massey
"An error bound for random tree codes".
Presented at the IEEE Int. Symp. Information Theory,
Ashkelon, Israel, June 25-29, 1973
- [7] R.G. Gallager
"Information Theory and Reliable Communication".
New York: Wiley, 1968, p 285
- [8] F. Jelinek
"A fast sequential decoding algorithm using a stack".
IBM J. Res. Dev., vol. 13, November 1969, 675-685
- [9] K.Sh. Zigangirov
"Some sequential decoding procedures".
Probl. Peredach. Inform., vol 2, 1966, No 4, 13-25

TABLE I

Channel transition probability	R_0	R/R_0	Metric		
			Code symbol	Received symbol	
				0	1
$p = 0.033$	0.50	0.89	0	1	-10
			1	-10	1
$p = 0.045$	0.50	1.00	0	1	-9
			1	-9	1
$p = 0.057$	0.45	1.11	0	1	-9
			1	-9	1

TABLE II

The probability of computational overflow P_o for the sequential decoding simulations. (1000 decoded frames).

p	M \ L	8	16	32	64	96	128	160	192	224	256	288	320
.033	4	.000	.000	.000	.000	.001	.001	.001	.001	.001	.001	.002	.002
	6	.000	.000	.000	.000	.001	.002	.002	.004	.008	.009	.009	.009
	8	.000	.000	.000	.000	.001	.001	.002	.004	.005	.006	.007	.008
	12	.000	.000	.000	.000	.001	.001	.001	.004	.006	.008	.008	.009
	16	.000	.000	.000	.001	.002	.005	.005	.008	.010	.013	.014	.016
.045	4	.000	.000	.000	.001	.003	.004	.005	.009	.013	.016	.019	.026
	6	.000	.000	.000	.002	.002	.009	.012	.022	.030	.036	.038	.049
	8	.000	.000	.000	.002	.006	.013	.016	.034	.044	.052	.057	.071
	12	.000	.000	.000	.002	.005	.012	.020	.031	.042	.050	.058	.073
	16	.000	.000	.000	.002	.006	.012	.018	.034	.043	.051	.059	.073
.057	4	.000	.000	.000	.006	.015	.021	.031	.052	.067	.086	.106	.127
	6	.000	.000	.000	.009	.029	.048	.070	.106	.133	.153	.176	.218
	8	.000	.000	.000	.007	.030	.051	.071	.130	.161	.189	.216	.249
	12	.000	.000	.000	.007	.026	.057	.078	.123	.152	.180	.215	.255
	16	.000	.000	.000	.008	.030	.055	.084	.132	.160	.190	.225	.277

TABLE III

Results of simulations at $p = 0.033$ ($R = 0.9R_0$)
(1000 decoded frames).

M	L	8	16	32	64	96	128	160	192	224	256	288	320
4	c	.075	.113	.113	.100	.116	.131	.134	.134	.117	.144	.137	.153
	R_T	.69	.62	.64	.51	.43	.52	.36	.33	.32	.38	.28	.31
6	c	.073	.111	.106	.076	.092	.099	.095	.102	.064	.079	.095	.106
	R_T	.97	.75	1.18	.52	.58	.51	.64	.57	.55	.57	.56	.54
8	c	.073	.112	.110	.094	.093	.114	.096	.104	.070	.078	.092	.097
	R_T	.97	.70	.99	.62	.59	.71	.62	.58	.54	.51	.53	.52
12	c	.061	.102	.106	.083	.098	.102	.095	.080	.058	.090	.088	.091
	R_T	.65	.60	1.29	.61	.58	.65	.63	.55	.65	.69	.66	.54
16	c	.073	.108	.106	.081	.096	.103	.087	.078	.073	.093	.092	.090
	R_T	.75	.65	.98	.68	.70	.72	.58	.58	.66	.77	.63	.64

TABLE IV

Results of simulation at $p = 0.045$ ($R = R_0$)
(1000 decoded frames).

M	L	8	16	32	64	96	128	160	192	224	256	288	320
4	c	.127	.180	.172	.199	.208	.234	.252	.276	.266	.296	.305	.344
	R_T	.52	.50	.45	.36	.32	.27	.21	.16	.13	.16	.13	.14
6	c	.101	.159	.158	.154	.159	.172	.167	.169	.141	.168	.170	.179
	R_T	.57	.50	.50	.50	.46	.48	.50	.41	.41	.46	.33	.39
8	c	.106	.156	.175	.168	.153	.154	.170	.154	.127	.132	.146	.139
	R_T	.60	.54	.47	.48	.47	.45	.45	.48	.50	.50	.43	.49
12	c	.099	.157	.163	.158	.154	.164	.158	.154	.104	.141	.145	.140
	R_T	.53	.51	.58	.49	.47	.53	.49	.46	.51	.51	.46	.48
16	c	.103	.151	.153	.149	.143	.152	.162	.140	.111	.138	.145	.146
	R_T	.48	.48	.47	.54	.50	.51	.50	.48	.49	.61	.50	.52

TABLE V

Results of simulations at $p = 0.057$ ($R = 1.1R_0$)
(1000 decoded frames).

M	L	8	16	32	64	96	128	160	192	224	256	288	320
4	c	.172	.227	.246	.266	.308	.367	.387	.406	.406	.424	.430	.459
	R_T	.48	.41	.38	.27	.20	.18	.15	.11	.10	.10	.07	.08
6	c	.153	.200	.226	.225	.216	.259	.242	.246	.214	.237	.244	.231
	R_T	.51	.37	.42	.39	.31	.36	.32	.28	.24	.29	.23	.25
8	c	.133	.207	.232	.244	.237	.260	.242	.219	.189	.177	.193	.187
	R_T	.42	.40	.38	.39	.35	.37	.37	.36	.35	.36	.30	.38
12	c	.139	.196	.223	.244	.229	.250	.233	.206	.168	.177	.179	.172
	R_T	.45	.38	.43	.42	.36	.41	.37	.36	.41	.37	.36	.42
16	c	.142	.204	.229	.225	.225	.251	.222	.194	.152	.188	.183	.162
	R_T	.49	.42	.40	.45	.33	.48	.41	.37	.39	.45	.39	.42

TABLE VI

Channel transition probability	\overline{R}_T	$E_{VU}(R)$	$E_{VL}(R)$
$p = 0.033$	0.63	0.56	0.68
$p = 0.045$	0.50	0.50	0.50
$p = 0.057$	0.39	0.38	0.38

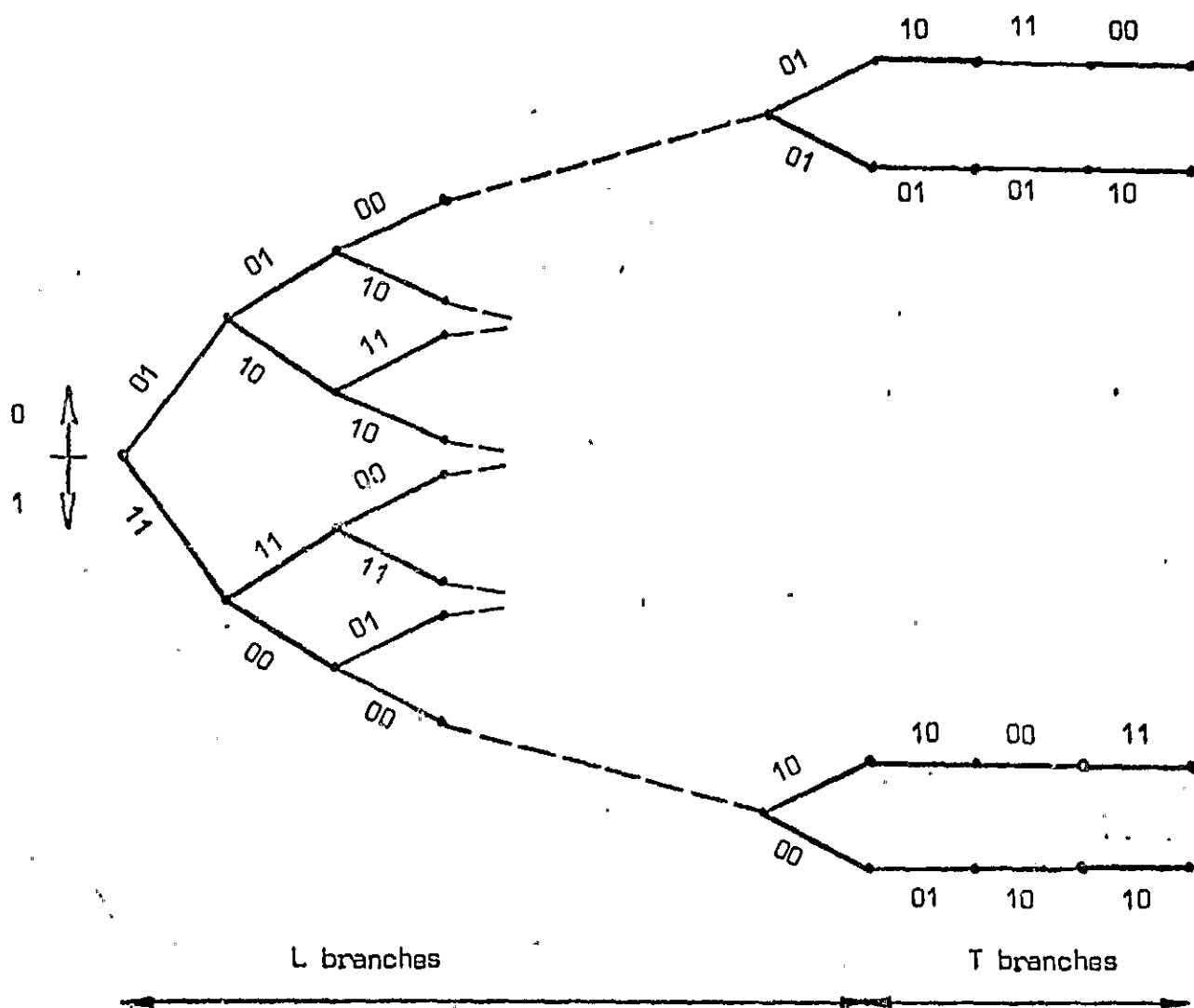
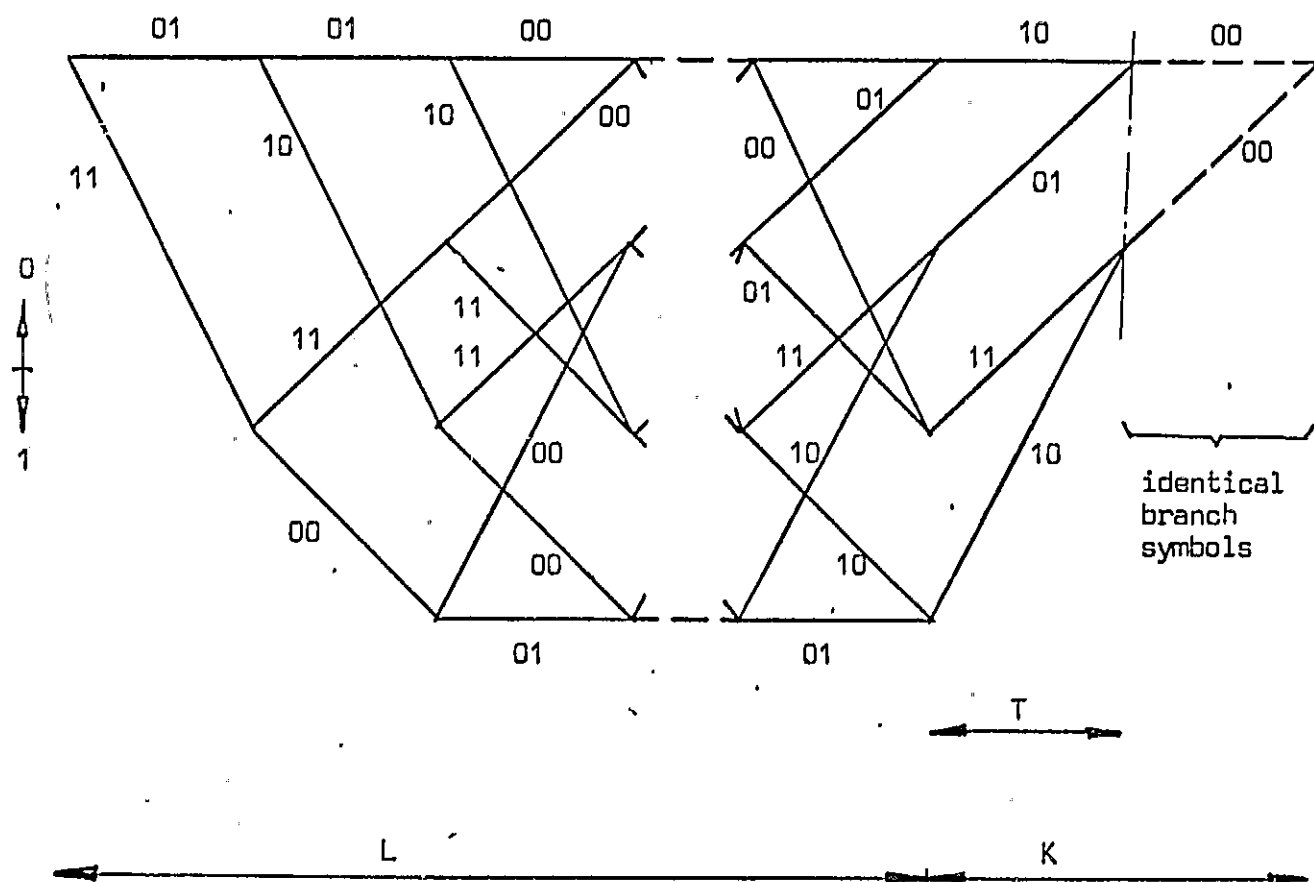


Fig. 1.

An example of a binary tree code with rate $1/2$ and tail length 3 for a binary input channel.



L, K and T in terms of branches.

Fig. 2.

An example of a binary trellis code with rate $1/2$, memory length 2 and tail length 1 for a binary input channel.

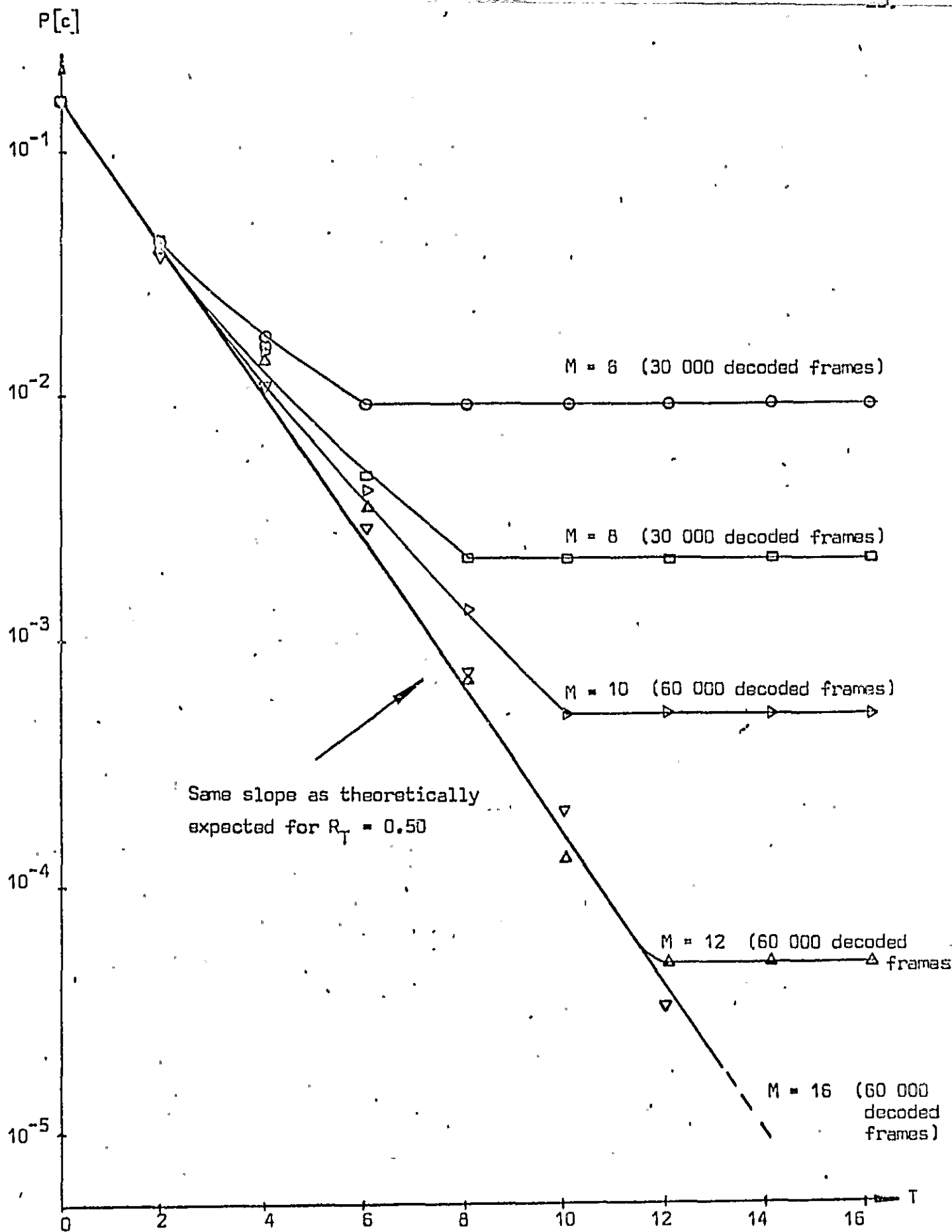


Fig. 3.

The decoding error probability obtained from sequential decoding simulations versus the tail length of the convolutional code.

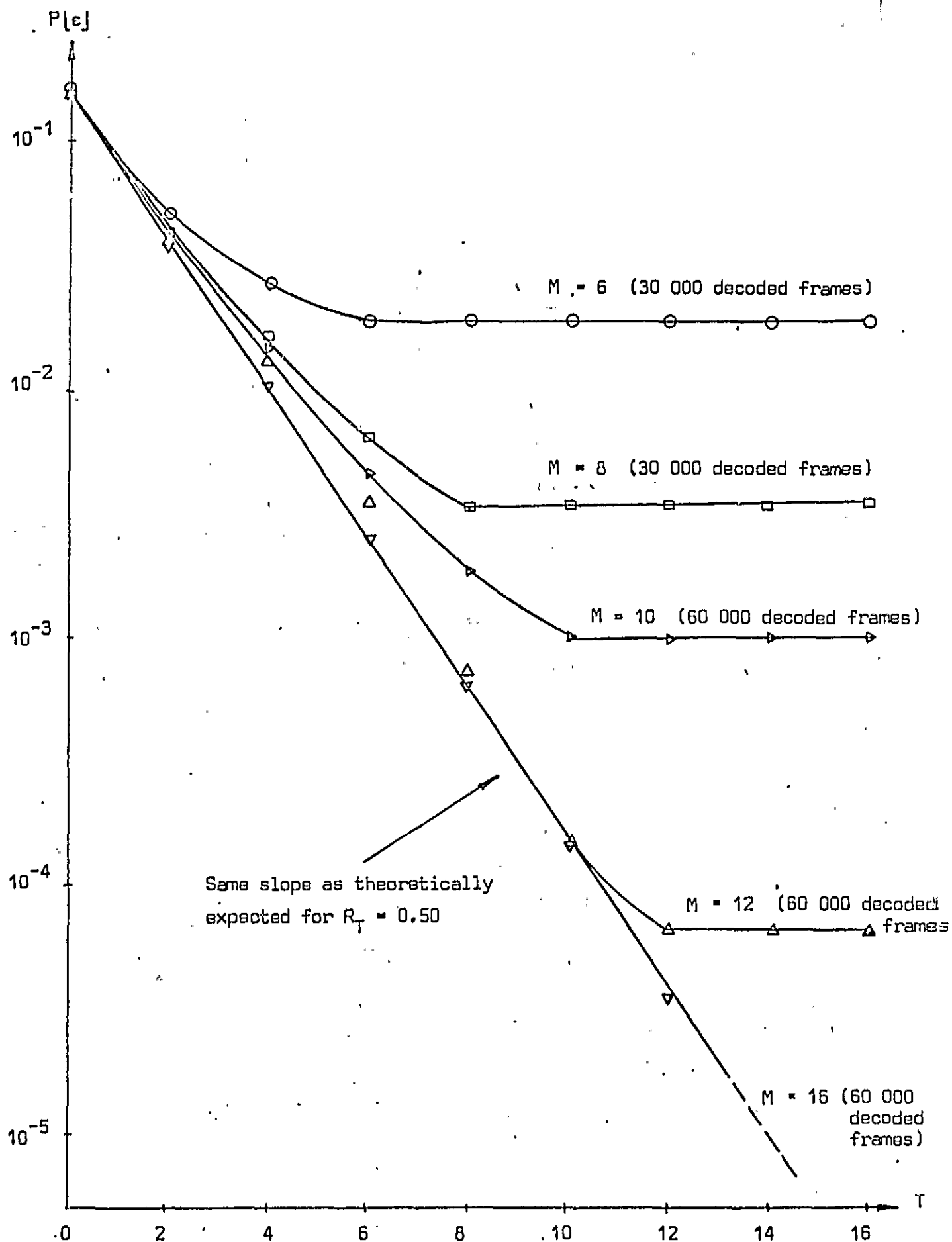


Fig. 4.

The decoding error probability obtained from sequential decoding simulations versus the tail length of the convolutional code. ($L = 128$, $p = 0.04$)

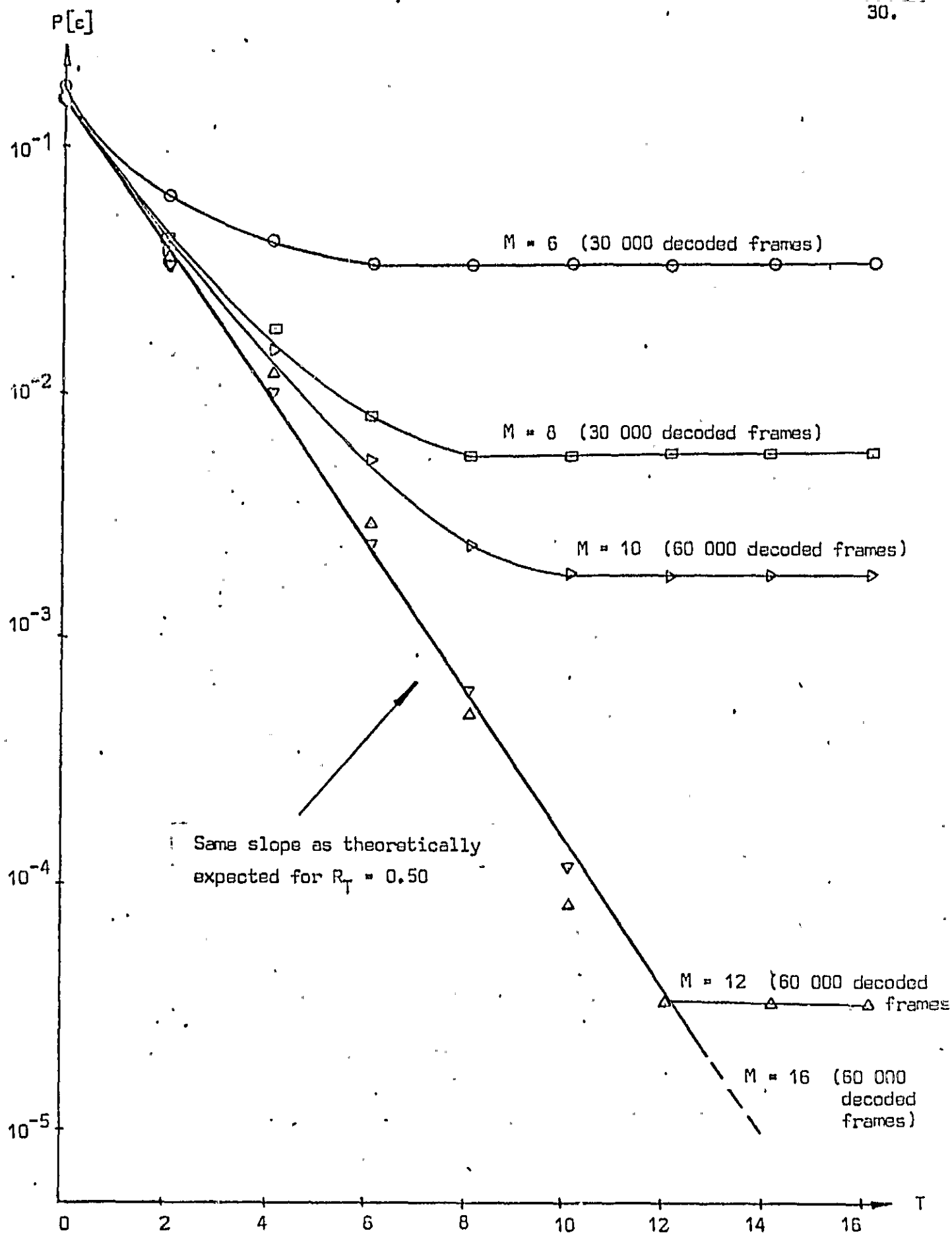


Fig. 5.

The decoding error probability obtained from sequential decoding simulations versus the tail length of the convolutional code.

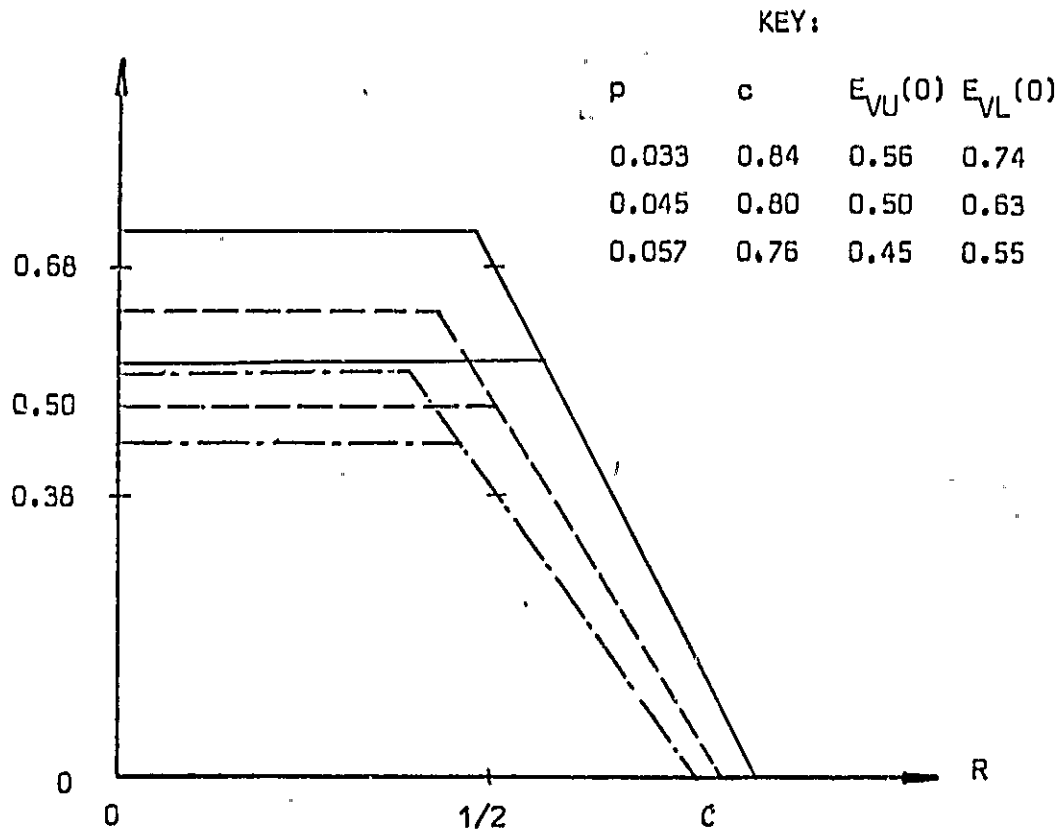


Fig. 6

Viterbi's exponents $E_{VU}(R)$ and $E_{VL}(R)$ for several binary symmetric channels

($p = 0.033$, $p = 0.045$, $p = 0.057$).